Article

Not peer-reviewed version

Learnable Entire-Kernel Spectral Layers: Discrete Frames, De-Aliasing, BM Regularization, and Beyond-Density MIT Injectivity

Mohammad Abu-Ghuwaleh * and Samir Brahim Belhaouari *

Posted Date: 16 March 2026

doi: 10.20944/preprints202603.1171.v1

Keywords: entire kernel spectral transforms; Master Integral Transform (MIT); Tight frames; spectral feature layers; de-aliasing algorithms; Beurling–Malliavin density; bandlimited reconstruction; nonharmonic spectral systems; signal inversion; computational harmonic analysis

 doi:10.20944/preprints202603.1171.v1

Article

Learnable Entire-Kernel Spectral Layers: Discrete Frames, De-Aliasing, BM Regularization, and Beyond-Density MIT Injectivity

Mohammad Abu-Ghuwaleh [1,*] and Samir Brahim Belhaouari [2,*]

[1] Department of Mathematics, Faculty of Science, Zarqa University, Zarqa 13110, Jordan
[2] Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
[*] Correspondence: 20209030@zu.edu.jo (M.A.-G.); sbelhaouari@hbku.edu.qa (S.B.B.)

Abstract

We study learnable spectral layers whose feature family is generated by an entire function, motivated by the Master Integral Transform (MIT) of [1]. In a periodic discrete model on $\mathbb{T}$, we define an oversampled multi-β analysis operator $\mathcal{A}_{g,K}^{(M)}$ built from an entire generator g and show that, on the one-sided bandlimited subspace $\mathrm{PW}_K(\mathbb{T})$, it is a tight frame with an explicit inverse, sharp frame bounds, and noise-stability constants governed solely by the Taylor coefficients of g. For general (non-bandlimited) signals we derive exact aliasing identities and quantify two de-aliasing mechanisms: a deterministic multi-resolution cancellation scheme and a randomized rotation estimator with unbiasedness, $\mathrm{MSE} = O(1/m)$, and high-probability bounds. We extend the discrete theory to $\mathbb{T}^d$, allowing general multivariate entire generators $G(z) = \sum_{\alpha\in\mathbb{N}^d} c_\alpha z^\alpha$, and obtain exact inversion and conditioning bounds on tensor bands $\mathrm{PW}_K^{(d)}(\mathbb{T}^d)$ with explicit constants. To connect discrete layers to continuum MIT injectivity, we formalize density control of active Taylor indices. We prove that bounded gaps imply positive one-sided interior Beurling–Malliavin density (hence, in particular, positive lower counting density), closing an end-to-end bridge from gap-regularized learning to the injectivity theorem of [1]. For bounded-gap sequences we also give a weighted-series characterization of strong a-regularity, yielding computable surrogate penalties. Finally, we prove two injectivity mechanisms that do not rely on density: (i) a β-analytic injectivity theorem (access to multiple β-channels near 0) for any nonconstant entire kernel, and (ii) a finite-band generic-shift result ensuring invertibility on PW_K for nonpolynomial generators. Full-data experiments illustrate conditioning collapse without coefficient floors and confirm the predicted $1/m$ de-aliasing variance decay. **Contributions (take-home).** (i) *Certified band-invertible spectral layers:* exact inversion + frame bounds + noise stability on $\mathrm{PW}_K(\mathbb{T})$ and $\mathrm{PW}_K^{(d)}(\mathbb{T}^d)$ with constants controlled by Taylor coefficients; (ii) *Provable de-aliasing:* deterministic multi-resolution cancellation and randomized rotation Monte Carlo with unbiasedness, $\mathrm{MSE} = O(1/m)$, and high-probability bounds; (iii) *A closed discrete-to-continuum bridge:* bounded-gap activity $\Rightarrow D_{BM} > 0$ and hence the lower counting density required by MIT injectivity; (iv) *Beyond-density injectivity + nonharmonic structure:* two density-free injectivity mechanisms and a monomial rigidity principle. **Applications.** These results yield *MIT-certified* modules for machine learning and scientific computing: invertible FFT-like embeddings for grid data, learnable positional encodings, and drop-in replacements for Fourier blocks in neural operators with explicit conditioning control and provable de-aliasing.

Keywords: entire kernel spectral transforms; Master Integral Transform (MIT); Tight frames; spectral feature layers; de-aliasing algorithms; Beurling–Malliavin density; bandlimited reconstruction; nonharmonic spectral systems; signal inversion; computational harmonic analysis

1. Introduction

Spectral representations are a foundational tool across modern machine learning: random Fourier features approximate kernels, FFT blocks accelerate convolutions, and neural operators parameterize integral kernels in frequency space. Yet most deployed spectral layers rely on *fixed* harmonic bases, typically sinusoids. This rigidity can be suboptimal for nonstationary data, boundary layers, heterogeneous coefficients, and mixed regimes.

This paper proposes a different axis:

> *Learn the spectral kernel generator itself, but enforce analytic structure and density constraints that prevent degeneracy and preserve invertibility.*

The Master Integral Transform (MIT) introduced in [1] supports kernels generated by arbitrary entire functions of finite order, and links completeness/injectivity to density properties of the nonzero Taylor indices of the generator. In the specific MIT injectivity statement we invoke below (Theorem 7.1), the explicit hypothesis is a positive lower *counting* density $\rho_* > 0$ of these indices; bounded-gap (syndetic) activity implies this, and in fact also implies positive interior Beurling–Malliavin density as a stronger regularity. This suggests a new ML primitive: a *learnable entire-kernel transform layer* stabilized by *density control*.

1.1. Related Work

Fourier-based embeddings are ubiquitous in ML: *random Fourier features* approximate shift-invariant kernels by sampling frequencies and embedding inputs into a sinusoidal feature space [7], and closely related sinusoidal features are widely used as fixed positional encodings. Neural operators such as the *Fourier Neural Operator* (FNO) learn mappings between function spaces by truncating and mixing Fourier modes [8]; this is powerful, but the backbone basis remains the standard harmonic one. A different thread seeks *invertibility* at the network level: normalizing flows (e.g. NICE/RealNVP/Glow) and invertible residual constructions enforce bijectivity (or controlled Jacobians) by architectural constraints [9–13]. Our angle is orthogonal: we keep the spectral viewpoint but enlarge the kernel class from fixed harmonics to *learnable entire-kernel families*, and we provide explicit analytic criteria (coefficient floors on finite bands; density constraints in the continuum) that prevent representational collapse and support explicit inversion.

1.2. Main Contributions

The manuscript is written to keep the discrete theory proof-complete and constant-explicit, while making the continuum assumptions transparent. The main contributions are:

- **Discrete periodic theory on $\mathbb{T}$ and $\mathbb{T}^d$ (proved).** For entire generators we construct oversampled multi-β analysis operators whose restriction to one-sided bandlimited subspaces is a tight frame. We give exact inversion formulae, sharp frame bounds, and explicit noise-stability constants, and we extend all results to $\mathbb{T}^d$ for general (not necessarily separable) multivariate entire generators.
- **Aliasing identities and quantitative de-aliasing (proved).** For general signals we derive exact residue-class aliasing formulae and analyze two de-aliasing mechanisms: a deterministic multi-resolution cancellation scheme and a randomized rotation estimator with unbiasedness, MSE $=$ $O(1/m)$, and high-probability bounds.
- **Density control and the continuum bridge to MIT injectivity (proved).** We formalize *BM regularization* of Taylor-index activity and prove that bounded gaps imply positive one-sided interior Beurling–Malliavin density with an explicit lower bound. In particular this forces the positive lower counting density condition required in the MIT injectivity theorem of [1]. For bounded-gap sequences we also provide a weighted-series characterization of strong a-regularity, yielding computable surrogate penalties.
- **Beyond-density injectivity and nonharmonic systems (proved).** We show two injectivity mechanisms that do not rely on density (a β-analytic theorem and a finite-band generic-shift theorem)

 doi:10.20944/preprints202603.1171.v1

and we isolate a rigidity phenomenon: except for monomials, entire generators produce nonharmonic orbit systems that cannot be reduced to single-frequency Fourier exponentials.

Contribution map (what you can actually *do* with the theorems).

To make the "theory $\to$ implementation" pathway explicit, Table 1 summarizes the main proven statements and the concrete engineering/AI functionality they unlock.

Table 1. Where the contributions live, and why they matter for applications.

Label	Core theorem(s)	What it enables in practice
(C1)	Theorems 3.3, 3.5, 3.6	A *certified*, band-invertible analysis/synthesis pair on $\mathrm{PW}_K(\mathbb{T})$ with explicit conditioning and noise amplification constants.
(C2)	Theorems 4.2, 4.3, 4.6, 4.7	Two provable de-aliasing modes for non-bandlimited signals: (i) multi-resolution cancellation (exact when the tail is finitely supported), and (ii) Monte Carlo random rotations with unbiasedness + $1/m$ variance + high-probability error control.
(C3)	Theorems 5.3, 5.4, 5.5	A full $\mathbb{T}^d$ version with *general* multivariate entire generators (not only separable/tensor ones), yielding certified inversion and de-aliasing on grid data.
(C4)	Theorems 6.3, 7.3	A closed "learning regularizer $\to$ analytic injectivity" bridge: bounded gaps $\Rightarrow D_{BM} > 0$ and (in particular) the lower counting density required by the MIT injectivity theorem.
(C5)	Theorems 8.1, 8.3	Two density-free injectivity mechanisms that justify additional experimental regimes (multi-β channels, or generic shifts) where completeness holds without asymptotic density hypotheses.
(C6)	Theorems 9.2, 9.3	A structural description of the induced nonharmonic systems (Gram design cone + monomial rigidity), clarifying what is genuinely "beyond Fourier" in the atom family.

Application lens.

The guiding use-case is *not* "replace FFT because FFT is slow"—FFT is unbeatable at what it does. The use-case is: *replace fixed Fourier atoms by a <u>learnable</u> analytic generator while keeping the one property engineers actually need: certified invertibility and controlled conditioning on the band you care about, plus principled de-aliasing when you leave the band.*

1.3. Structure

Section 2 sets notation and recalls the density notions used (counting density and one-sided interior BM density). Section 3 develops the complete $\mathbb{T}$-theory: exact inversion on $\mathrm{PW}_K(\mathbb{T})$, tight frame bounds, and noise stability, followed by an aliasing identity for general signals. Section 4 proves two de-aliasing mechanisms (multi-resolution cancellation and random-rotation Monte Carlo) with quantitative guarantees. Section 5 extends the full theory to $\mathbb{T}^d$ for general multivariate entire generators. Section 6 formalizes BM regularization and proves "bounded gaps $\Rightarrow$ positive interior BM density"; Section 6.4 adds a bounded-gap series characterization of strong regularity. Section 7 closes the discrete-to-continuum bridge by verifying the density hypotheses of the MIT injectivity theorem [1] under gap regularization. Section 8 records two injectivity mechanisms beyond density (a β-analytic theorem and a generic-shift theorem), and Section 9 isolates nonharmonic orbit structure and a monomial rigidity principle. Section 10 discusses AI-motivated architectures, and Section 11 provides full-data experiments. Section 12 concludes with remarks and open problems.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1171.v1

2. Preliminaries: Fourier, Entire Kernels, and BM Density

2.1. Periodic Domain and Bandlimited Subspaces

Let $\mathbb{T} = [0, 2\pi)$ with normalized measure $dx/(2\pi)$. For $f \in L^2(\mathbb{T})$ define Fourier coefficients

$$\hat{f}(k) := \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx}\, dx, \qquad k \in \mathbb{Z}.$$

Parseval gives $\|f\|^2_{L^2(\mathbb{T})} = \sum_{k\in\mathbb{Z}} |\hat{f}(k)|^2$.

Remark 2.1. *Our layers below use kernels $g(\beta e^{-ix}) = \sum_{k\geq 0} c_k \beta^k e^{-ikx}$, hence they naturally interact with the* nonnegative *Fourier indices $k \geq 0$. This is why we work with one-sided bandlimits $\{0, \dots, K\}$.*

Definition 2.2 (K-bandlimited (one-sided) subspace)**.** *For $K \in \mathbb{N}$ define*

$$\mathrm{PW}_K(\mathbb{T}) := \left\{ f(x) = \sum_{k=0}^{K} a_k e^{ikx} \ : \ a_k \in \mathbb{C} \right\}.$$

Equivalently, $f \in \mathrm{PW}_K(\mathbb{T})$ iff $\hat{f}(k) = 0$ for all $k \notin \{0, 1, \dots, K\}$.

2.2. Entire Kernel Generators

Let $g : \mathbb{C} \to \mathbb{C}$ be entire with Taylor expansion about 0:

$$g(z) = \sum_{k=0}^{\infty} c_k z^k, \qquad c_k \in \mathbb{C}. \tag{1}$$

(Any polynomial is an entire function of finite order; thus the discrete theory below is automatically within the MIT admissible kernel class.)

Definition 2.3 (Active index set)**.** $S(g) := \{k \in \mathbb{N} : \ c_k \neq 0\}$.

Lemma 2.4 (Kernel expansion)**.** *For any $\beta \in \mathbb{C}$ and $x \in \mathbb{R}$,*

$$g(\beta e^{-ix}) = \sum_{k=0}^{\infty} c_k \beta^k e^{-ikx},$$

with absolute convergence for each fixed x.

Proof. Substitute $z = \beta e^{-ix}$ into (1). □

2.3. Interior Beurling–Malliavin Density and Strong Regularity

We use a standard definition of interior BM density via regular subsequences.

Definition 2.5 (Counting function and strong a-regularity)**.** *Let $\Lambda \subset \mathbb{R}_{\geq 0}$ be a discrete set. Define the counting function*

$$n_\Lambda(x) := \#\{\lambda \in \Lambda : \ \lambda \leq x\}, \qquad x \geq 0.$$

We say Λ is strongly a-regular *(for $a \geq 0$) if*

$$\int_0^{\infty} \frac{|n_\Lambda(x) - ax|}{1 + x^2}\, dx < \infty. \tag{2}$$

Remark 2.6. *Strong a-regularity is a standard proxy for BM regularity; see, e.g., [2], which notes that "a-regular" and "strongly a-regular" lead to equivalent density definitions in this context.*

Definition 2.7 (One-sided interior BM density)**.** *For a discrete $\Lambda \subset \mathbb{R}_{\geq 0}$ define its (one-sided)* interior BM density *by*

$$D_{BM}(\Lambda) \ := \ \sup\{a \geq 0 : \ \exists\, \Lambda_0 \subset \Lambda \text{ a subsequence that is strongly } a\text{-regular}\}.$$

 doi:10.20944/preprints202603.1171.v1

Remark 2.8. *Interior and exterior BM densities are classically defined on two-sided sequences in $\mathbb{R}$. Our use is one-sided because the MIT "Taylor index set" is naturally a subset of $\mathbb{N}$. One-sided variants appear in the MIT framework and in one-sided completeness theorems; see [1] and references therein. For a recent discussion relating Beurling–Malliavin density to asymptotic (counting) density, see [4].*

2.4. Syndetic Sets (Bounded Gaps)

Definition 2.9 (L-syndetic (bounded gaps)). *A set $S \subset \mathbb{N}$ is* L-syndetic *if every block of integers of length L contains at least one element of S, i.e.,*

$$\forall m \in \mathbb{N}, \quad S \cap \{mL, mL+1, \dots, (m+1)L-1\} \neq \emptyset.$$

3. Entire-Kernel Spectral Layers on $\mathbb{T}$: Frames and Exact Inversion

3.1. Oversampled Multi-β Analysis Operator

Fix integers $K \in \mathbb{N}$ and $M \in \mathbb{N}$ with $M \geq K+1$. Set $N := K+1$ and define

$$\omega_M := e^{2\pi i/M}, \qquad \beta_r := \omega_M^r, \qquad r = 0, 1, \dots, M-1.$$

Definition 3.1 ($\mathbb{T}$-MIT analysis operator (oversampled)). *Let g be entire. Define $\mathcal{A}_{g,K}^{(M)} : L^2(\mathbb{T}) \to \mathbb{C}^M$ by*

$$(\mathcal{A}_{g,K}^{(M)} f)_r := \frac{1}{2\pi} \int_0^{2\pi} f(x)\, g(\beta_r e^{-ix})\, dx, \qquad r = 0, \dots, M-1. \tag{3}$$

Proposition 3.2 (Fourier representation). *If $f \in \mathrm{PW}_K(\mathbb{T})$ then*

$$(\mathcal{A}_{g,K}^{(M)} f)_r = \sum_{k=0}^{K} c_k \beta_r^k \hat{f}(k), \qquad r = 0, \dots, M-1. \tag{4}$$

Proof. Insert Lemma 2.4 into (3) and use orthogonality of exponentials. Terms with $k > K$ vanish since $f \in \mathrm{PW}_K$. □

3.2. Exact Inversion via Pseudoinverse (Tight Frame)

Define the bandwise coefficient extrema

$$m_K(g) := \min_{0 \leq k \leq K} |c_k|, \qquad M_K(g) := \max_{0 \leq k \leq K} |c_k|.$$

Theorem 3.3 (Exact inversion for $M \geq K+1$). *Assume $c_k \neq 0$ for $k = 0, \dots, K$. For $f \in \mathrm{PW}_K(\mathbb{T})$ define $y = \mathcal{A}_{g,K}^{(M)} f \in \mathbb{C}^M$. Then*

$$\hat{f}(k) = \frac{1}{c_k} \cdot \frac{1}{M} \sum_{r=0}^{M-1} y_r\, \beta_r^{-k}, \qquad k = 0, \dots, K, \tag{5}$$

and $f(x) = \sum_{k=0}^{K} \hat{f}(k) e^{ikx}$.

Proof. For $k, \ell \in \{0, \dots, K\}$ with $K < M$, orthogonality of roots of unity yields

$$\frac{1}{M} \sum_{r=0}^{M-1} \beta_r^{k-\ell} = \delta_{k\ell}.$$

Multiply (4) by β_r^{-k}, sum over r, divide by M, and isolate $\hat{f}(k)$. □

Remark 3.4. *When $M = K+1$, (5) is an inverse DFT (square case). When $M > K+1$, it is the Moore–Penrose pseudoinverse of a partial Fourier matrix; the columns remain orthogonal, hence the formula is still exact and numerically stable.*

 doi:10.20944/preprints202603.1171.v1

3.3. Sharp Frame Bounds and Conditioning

Theorem 3.5 (Tight frame bounds)**.** *Assume $m_K(g) > 0$. Then for all $f \in \mathrm{PW}_K(\mathbb{T})$,*

$$\sqrt{M}\, m_K(g)\, \|f\|_{L^2(\mathbb{T})} \;\le\; \|\mathcal{A}_{g,K}^{(M)} f\|_{\ell^2} \;\le\; \sqrt{M}\, M_K(g)\, \|f\|_{L^2(\mathbb{T})}. \tag{6}$$

Consequently the analysis map is bi-Lipschitz on $\mathrm{PW}_K(\mathbb{T})$.

Proof. Let $\hat{f}_{0:K} \in \mathbb{C}^{K+1}$ be the coefficient vector. Write $y = V_{M,K} D_g \hat{f}_{0:K}$ where $(V_{M,K})_{r,k} = \beta_r^k$ and $D_g = \mathrm{diag}(c_0, \dots, c_K)$. Because $K < M$, the columns of $V_{M,K}$ are orthogonal with $\|V_{M,K} e_k\|_2 = \sqrt{M}$. Thus $\|V_{M,K} u\|_2 = \sqrt{M}\|u\|_2$ for all $u \in \mathbb{C}^{K+1}$, hence

$$\|y\|_2 = \sqrt{M}\, \|D_g \hat{f}_{0:K}\|_2.$$

Bound $\|D_g \hat{f}\|_2$ above and below by $M_K(g)\|\hat{f}\|_2$ and $m_K(g)\|\hat{f}\|_2$, and use Parseval $\|\hat{f}\|_2 = \|f\|_{L^2(\mathbb{T})}$ on PW_K. □

3.4. Noise Stability

Theorem 3.6 (Stable inversion under additive feature noise)**.** *Assume $m_K(g) > 0$ and let $f \in \mathrm{PW}_K(\mathbb{T})$. Suppose we observe*

$$y = \mathcal{A}_{g,K}^{(M)} f + \eta, \qquad \eta \in \mathbb{C}^M.$$

Let $\tilde{f}$ be reconstructed from y by (5). Then

$$\|\tilde{f} - f\|_{L^2(\mathbb{T})} \le \frac{1}{\sqrt{M}\, m_K(g)}\, \|\eta\|_{\ell^2}. \tag{7}$$

Proof. Let $\delta y = \eta$ and $\delta \hat{f}$ be the coefficient error. From (5),

$$\delta \hat{f}(k) = \frac{1}{c_k} \cdot \frac{1}{M} \sum_{r=0}^{M-1} \eta_r \beta_r^{-k}.$$

Let $V_{M,K}^*$ denote the adjoint partial Fourier matrix. Then $\delta \hat{f} = D_g^{-1} \frac{1}{M} V_{M,K}^* \eta$. Since $\|V_{M,K}^* \eta\|_2 \le \|V_{M,K}\|\, \|\eta\|_2 = \sqrt{M}\|\eta\|_2$ and $\|D_g^{-1}\| \le 1/m_K(g)$,

$$\|\delta \hat{f}\|_2 \le \frac{1}{m_K(g)} \cdot \frac{1}{M} \cdot \sqrt{M}\, \|\eta\|_2 = \frac{1}{\sqrt{M}\, m_K(g)} \|\eta\|_2.$$

Finally $\|\tilde{f} - f\|_{L^2} = \|\delta \hat{f}\|_2$ by Parseval on PW_K. □

4. Aliasing and De-Aliasing on $\mathbb{T}$

4.1. Exact Aliasing Identity for General Signals

For general $f \in L^2(\mathbb{T})$, the oversampled inverse-DFT statistic

$$s_k^{(M)}(f) := \frac{1}{M} \sum_{r=0}^{M-1} (\mathcal{A}_{g,K}^{(M)} f)_r\, \beta_r^{-k}, \qquad k = 0, \dots, K, \tag{8}$$

isolates residue classes modulo M for the nonnegative Fourier indices.

Lemma 4.1 (Residue selection)**.** *For any $k \in \{0, \dots, K\}$,*

$$s_k^{(M)}(f) = \sum_{\ell \ge 0} c_{k+\ell M}\, \hat{f}(k + \ell M). \tag{9}$$

 doi:10.20944/preprints202603.1171.v1

Proof. Expand $(\mathcal{A}^{(M)}_{g,K} f)_r = \sum_{n\geq 0} c_n \beta_r^n \hat{f}(n)$ (by Lemma 2.4 and orthogonality). Then

$$s_k^{(M)}(f) = \sum_{n\geq 0} c_n \hat{f}(n) \left(\frac{1}{M} \sum_{r=0}^{M-1} \beta_r^{n-k} \right),$$

and the roots-of-unity sum equals 1 iff $n \equiv k \pmod M$, else 0. □

Theorem 4.2 (Aliasing identity and tail bound)**.** *Assume $c_k \neq 0$ for $k \leq K$. Define the naive coefficient estimator*

$$\widetilde{\hat{f}}^{(M)}(k) := \frac{1}{c_k} s_k^{(M)}(f).$$

Then

$$\widetilde{\hat{f}}^{(M)}(k) = \hat{f}(k) + \frac{1}{c_k} \sum_{\ell \geq 1} c_{k+\ell M} \hat{f}(k+\ell M). \tag{10}$$

Consequently, defining $\widetilde{f}(x) = \sum_{k=0}^{K} \widetilde{\hat{f}}^{(M)}(k) e^{ikx}$ and P_K the K-projection,

$$\|\widetilde{f} - P_K f\|^2_{L^2(\mathbb{T})} \leq \frac{1}{m_K(g)^2} \sum_{k=0}^{K} \left(\sum_{\ell\geq 1} |c_{k+\ell M}|\, |\hat{f}(k+\ell M)| \right)^2. \tag{11}$$

Proof. Identity (10) is Lemma 4.1 divided by c_k. The bound uses $|c_k|^{-1} \leq 1/m_K(g)$ and Parseval on PW_K. □

4.2. Multi-Resolution Cancellation (Multi-Stage)

The aliasing identity shows that low coefficients are contaminated by a structured arithmetic-progression tail. Multi-resolution cancellation uses a larger band to estimate and subtract leading aliases.

Theorem 4.3 (Two-band alias cancellation: exact for finite support)**.** *Let $K_2 \geq K_1$ and choose $M_1 \geq K_1 + 1$, $M_2 \geq K_2 + 1$. Assume $c_k \neq 0$ for $0 \leq k \leq K_2$. Let $f \in \mathrm{PW}_{K_2}(\mathbb{T})$. For each $k \in \{0, \ldots, K_1\}$ define*

$$\widehat{\hat{f}}(k) := \frac{1}{c_k} \left(s_k^{(M_1)}(f) - \sum_{\ell=1}^{L_k} c_{k+\ell M_1} \hat{f}(k+\ell M_1) \right), \qquad L_k := \left\lfloor \frac{K_2 - k}{M_1} \right\rfloor. \tag{12}$$

Then $\widehat{\hat{f}}(k) = \hat{f}(k)$ for all $k \leq K_1$. Moreover, replacing the true $\hat{f}(k+\ell M_1)$ in (12) by the fine-band reconstruction from $y^{(2)} = \mathcal{A}^{(M_2)}_{g,K_2} f$ yields an explicit exact reconstruction procedure.

Proof. Because $f \in \mathrm{PW}_{K_2}$, the residue sum (9) truncates at $\ell \leq L_k$. Thus $s_k^{(M_1)}(f) = c_k \hat{f}(k) + \sum_{\ell=1}^{L_k} c_{k+\ell M_1} \hat{f}(k+\ell M_1)$. Subtract the tail terms and divide by c_k. The remaining statement follows because the fine-band inversion is exact on PW_{K_2} (Theorem 3.3). □

Theorem 4.4 (Multi-stage cancellation (inductive exactness))**.** *Let $K_1 < K_2 < \cdots < K_T$ and choose $M_t \geq K_t + 1$. Assume $c_k \neq 0$ for $0 \leq k \leq K_T$ and let $f \in \mathrm{PW}_{K_T}(\mathbb{T})$. Define estimators for coefficients on $\{0, \ldots, K_1\}$ by iteratively cancelling aliases using reconstructions from successively finer bands (starting from band K_T and descending). Then the resulting coefficients are exact: $\widehat{\hat{f}}(k) = \hat{f}(k)$ for all $k \leq K_1$.*

Proof. Proceed by descending induction on the stages. At the finest stage T, inversion is exact on PW_{K_T}. Assume all coefficients up to K_{t+1} are known exactly. Then for each $k \leq K_t$, the residue identity truncates at frequencies $\leq K_{t+1}$, so subtracting the known alias terms yields exact $\hat{f}(k)$. Iterate down to $t = 1$. □

 doi:10.20944/preprints202603.1171.v1

4.3. Random-Rotation Monte Carlo De-Aliasing

Define a rotated β-grid by $a \in \mathbb{T}$:

$$\beta_r^{(a)} := a\,\beta_r.$$

Let $\mathcal{A}_{g,K}^{(M,a)}$ be the analysis operator (3) with β_r replaced by $\beta_r^{(a)}$.

Lemma 4.5 (Rotated alias identity). *Assume $c_k \neq 0$ for $k \leq K$. Define the rotated estimator*

$$\tilde{\hat{f}}_a^{(M)}(k) := \frac{1}{c_k\,a^k} \cdot \frac{1}{M} \sum_{r=0}^{M-1} (\mathcal{A}_{g,K}^{(M,a)} f)_r\, \beta_r^{-k}.$$

Then

$$\tilde{\hat{f}}_a^{(M)}(k) = \hat{f}(k) + \frac{1}{c_k} \sum_{\ell \geq 1} c_{k+\ell M}\, a^{\ell M}\, \hat{f}(k+\ell M). \tag{13}$$

Proof. Same calculation as Lemma 4.1; the extra factor a^n survives and becomes $a^{\ell M}$ on the residue class. $\square$

Theorem 4.6 (Unbiasedness and MSE). *Assume $c_k \neq 0$ for $k \leq K$. Let $a_1, \dots, a_m$ be i.i.d. uniform on $\mathbb{T}$ and define*

$$\bar{\hat{f}}^{(M)}(k) := \frac{1}{m} \sum_{j=1}^{m} \tilde{\hat{f}}_{a_j}^{(M)}(k).$$

Then for each $k \leq K$:

(i) $\mathbb{E}\,\bar{\hat{f}}^{(M)}(k) = \hat{f}(k)$.

(ii)

$$\mathbb{E}\left|\bar{\hat{f}}^{(M)}(k) - \hat{f}(k)\right|^2 = \frac{1}{m|c_k|^2} \sum_{\ell \geq 1} |c_{k+\ell M}|^2\, |\hat{f}(k+\ell M)|^2. \tag{14}$$

Summing over $k \leq K$ yields

$$\mathbb{E}\left\|\bar{f}^{(M)} - P_K f\right\|_{L^2(\mathbb{T})}^2 \leq \frac{1}{m\, m_K(g)^2} \sum_{n > K} |c_n|^2\, |\hat{f}(n)|^2, \tag{15}$$

where $\bar{f}^{(M)}(x) = \sum_{k=0}^{K} \bar{\hat{f}}^{(M)}(k) e^{ikx}$.

Proof. From Lemma 4.5, the alias term has factors $a^{\ell M}$ with zero mean for $\ell \geq 1$, giving unbiasedness. For the MSE, independence yields a $1/m$ variance reduction and orthogonality $\mathbb{E}[a^{\ell M}\overline{a^{\ell' M}}] = 0$ unless $\ell = \ell'$. The global bound follows because each $n > K$ belongs to exactly one residue class modulo M. $\square$

Theorem 4.7 (High-probability bound (bounded alias energy)). *Assume $c_k \neq 0$ for $k \leq K$. Fix $k \leq K$ and define the alias magnitude bound*

$$B_k := \sum_{\ell \geq 1} \left|\frac{c_{k+\ell M}}{c_k}\right| |\hat{f}(k+\ell M)|.$$

Then each rotated sample satisfies $|\tilde{\hat{f}}_a^{(M)}(k) - \hat{f}(k)| \leq B_k$. Consequently, for $\delta \in (0,1)$,

$$\mathbb{P}\left(\left|\bar{\hat{f}}^{(M)}(k) - \hat{f}(k)\right| \geq 2B_k \sqrt{\frac{\log(4/\delta)}{m}}\right) \leq \delta. \tag{16}$$

Proof. From Lemma 4.5, each rotated sample error can be written as

$$Z_a := \tilde{\hat{f}}_a^{(M)}(k) - \hat{f}(k) = \sum_{\ell \geq 1} \frac{c_{k+\ell M}}{c_k} a^{\ell M} \hat{f}(k+\ell M),$$

so $|Z_a| \le B_k$ since $|a^{\ell M}| = 1$ and B_k is the absolute-sum bound.

Let $a_1, \dots, a_m$ be i.i.d. uniform on $\mathbb{T}$ and set $Z_j := Z_{a_j}$, so that

$$\bar{\hat{f}}^{(M)}(k) - \hat{f}(k) = \frac{1}{m} \sum_{j=1}^{m} Z_j.$$

Write $X_j := \Re Z_j$ and $Y_j := \Im Z_j$. Then $|X_j| \le B_k$ and $|Y_j| \le B_k$, and (by the same zero-mean argument used in Theorem 4.6) we have $\mathbb{E}X_j = \mathbb{E}Y_j = 0$. Hoeffding's inequality gives, for any $t > 0$,

$$\mathbb{P}\Big(\Big|\frac{1}{m}\sum_{j=1}^{m} X_j\Big| \ge t\Big) \le 2\exp\Big(-\frac{mt^2}{2B_k^2}\Big), \qquad \mathbb{P}\Big(\Big|\frac{1}{m}\sum_{j=1}^{m} Y_j\Big| \ge t\Big) \le 2\exp\Big(-\frac{mt^2}{2B_k^2}\Big).$$

Finally, $|u + iv| \ge s$ implies $|u| \ge s/\sqrt{2}$ or $|v| \ge s/\sqrt{2}$, hence by a union bound

$$\mathbb{P}\Big(\Big|\frac{1}{m}\sum_{j=1}^{m} Z_j\Big| \ge s\Big) \le 4\exp\Big(-\frac{ms^2}{4B_k^2}\Big).$$

Setting $s = 2B_k\sqrt{\log(4/\delta)/m}$ yields (16). $\square$

5. Full $\mathbb{T}^d$ Extension with Constants (General Multivariate Generators)

5.1. Fourier and Bandlimited Subspaces on $\mathbb{T}^d$

Let $\mathbb{T}^d = [0, 2\pi)^d$ with normalized measure $dx/(2\pi)^d$. For $f \in L^2(\mathbb{T}^d)$ define

$$\hat{f}(k) := \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} f(x)\, e^{-i\langle k,x\rangle}\, dx, \qquad k \in \mathbb{Z}^d.$$

Define the tensor bandlimit

$$\mathrm{PW}_K^{(d)}(\mathbb{T}^d) := \Bigg\{ f(x) = \sum_{k \in \{0,\dots,K\}^d} a_k e^{i\langle k,x\rangle} \Bigg\}.$$

5.2. General Multivariate Entire Generators

Let $G : \mathbb{C}^d \to \mathbb{C}$ be entire with multivariate power series

$$G(z) = \sum_{\alpha \in \mathbb{N}^d} c_\alpha z^\alpha, \qquad z^\alpha := \prod_{j=1}^{d} z_j^{\alpha_j}. \tag{17}$$

Define the active index set $S(G) := \{\alpha \in \mathbb{N}^d : \ c_\alpha \ne 0\}$.

Fix $K \in \mathbb{N}$ and $M \ge K + 1$ and define $\omega_M = e^{2\pi i/M}$. For $r = (r_1, \dots, r_d) \in \{0, \dots, M-1\}^d$, set

$$\beta_r := (\omega_M^{r_1}, \dots, \omega_M^{r_d}) \in \mathbb{T}^d.$$

Definition 5.1 ($\mathbb{T}^d$-MIT analysis operator)**.** *Define $\mathcal{A}_{G,K}^{(d,M)} : L^2(\mathbb{T}^d) \to \mathbb{C}^{M^d}$ by*

$$(\mathcal{A}_{G,K}^{(d,M)} f)_r := \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} f(x)\, G(\beta_r \odot e^{-ix})\, dx, \qquad r \in \{0, \dots, M-1\}^d, \tag{18}$$

where $\odot$ denotes componentwise product and $e^{-ix} = (e^{-ix_1}, \dots, e^{-ix_d})$.

Proposition 5.2 (Fourier representation on $\mathrm{PW}_K^{(d)}$)**.** *If $f \in \mathrm{PW}_K^{(d)}(\mathbb{T}^d)$, then for each r,*

$$(\mathcal{A}_{G,K}^{(d,M)} f)_r = \sum_{\alpha \in \{0,\dots,K\}^d} c_\alpha\, \beta_r^\alpha\, \hat{f}(\alpha), \qquad \beta_r^\alpha := \prod_{j=1}^{d} (\omega_M^{r_j})^{\alpha_j}. \tag{19}$$

Proof. Expand $G(\beta_r \odot e^{-ix}) = \sum_\alpha c_\alpha \beta_r^\alpha e^{-i\langle\alpha,x\rangle}$ and use orthogonality on $\mathbb{T}^d$. $\square$

Define band extrema

$$m_K^{(d)}(G) := \min_{\alpha\in\{0,\ldots,K\}^d} |c_\alpha|, \qquad M_K^{(d)}(G) := \max_{\alpha\in\{0,\ldots,K\}^d} |c_\alpha|.$$

Theorem 5.3 (Exact inversion on $\mathrm{PW}_K^{(d)}(\mathbb{T}^d)$)**.** *Assume $c_\alpha \in \mathbb{C}\setminus\{0\}$ for all $\alpha \in \{0,\ldots,K\}^d$. Let $f \in \mathrm{PW}_K^{(d)}(\mathbb{T}^d)$ and $y = \mathcal{A}_{G,K}^{(d,M)} f$. Then for each $\alpha \in \{0,\ldots,K\}^d$,*

$$\hat{f}(\alpha) = \frac{1}{c_\alpha} \cdot \frac{1}{M^d} \sum_{r\in\{0,\ldots,M-1\}^d} y_r\, \beta_r^{-\alpha}. \tag{20}$$

Proof. Because $K < M$, the d-dimensional roots-of-unity orthogonality holds:

$$\frac{1}{M^d} \sum_r \beta_r^{\alpha-\gamma} = \delta_{\alpha\gamma} \quad \text{for } \alpha, \gamma \in \{0,\ldots,K\}^d.$$

Multiply (19) by $\beta_r^{-\alpha}$, sum over r, divide by M^d, and isolate $\hat{f}(\alpha)$. $\square$

Theorem 5.4 (Frame bounds and noise stability on $\mathbb{T}^d$)**.** *Assume $m_K^{(d)}(G) > 0$. Then for all $f \in \mathrm{PW}_K^{(d)}(\mathbb{T}^d)$,*

$$M^{d/2}\, m_K^{(d)}(G)\, \|f\|_{L^2(\mathbb{T}^d)} \;\le\; \|\mathcal{A}_{G,K}^{(d,M)} f\|_{\ell^2} \;\le\; M^{d/2}\, M_K^{(d)}(G)\, \|f\|_{L^2(\mathbb{T}^d)}. \tag{21}$$

If $y = \mathcal{A}_{G,K}^{(d,M)} f + \eta$, reconstructing by (20) yields

$$\|\tilde{f} - f\|_{L^2(\mathbb{T}^d)} \le \frac{1}{M^{d/2}\, m_K^{(d)}(G)} \|\eta\|_{\ell^2}. \tag{22}$$

Proof. Same as Theorems 3.5 and 3.6, using unitarity of the d-dimensional partial Fourier system and Parseval on $\mathrm{PW}_K^{(d)}$. $\square$

5.3. Aliasing and Random-Rotation De-Aliasing on $\mathbb{T}^d$

For general f, residue classes are now modulo M in each coordinate. Define

$$s_\alpha^{(d,M)}(f) := \frac{1}{M^d} \sum_r (\mathcal{A}_{G,K}^{(d,M)} f)_r\, \beta_r^{-\alpha}.$$

Then

$$s_\alpha^{(d,M)}(f) = \sum_{\ell\in\mathbb{N}^d} c_{\alpha+M\ell}\, \hat{f}(\alpha + M\ell), \tag{23}$$

where $M\ell = (M\ell_1, \ldots, M\ell_d)$ and $\alpha + M\ell$ is componentwise.

Random rotations use independent phases $a = (a_1, \ldots, a_d) \in \mathbb{T}^d$ and rotated grids $\beta_r^{(a)} = (a_1\beta_{r_1}, \ldots, a_d\beta_{r_d})$. Then alias terms acquire factors $\prod_{j=1}^d a_j^{M\ell_j}$, whose expectation is zero for any nonzero ℓ.

Theorem 5.5 (Unbiased random-rotation de-aliasing on $\mathbb{T}^d$)**.** *Assume $c_\alpha \neq 0$ for $\alpha \in \{0,\ldots,K\}^d$. Let $a^{(1)}, \ldots, a^{(m)}$ be i.i.d. uniform on $\mathbb{T}^d$ and define averaged estimators of band coefficients via rotated analogues of (20). Then the estimators are unbiased and satisfy an MSE bound*

$$\mathbb{E}\left\|\overline{f}^{(d,M)} - P_K f\right\|_{L^2(\mathbb{T}^d)}^2 \le \frac{1}{m\, m_K^{(d)}(G)^2} \sum_{\gamma\notin\{0,\ldots,K\}^d} |c_\gamma|^2\, |\hat{f}(\gamma)|^2.$$

 doi:10.20944/preprints202603.1171.v1

Proof. Expand as in (23). Nontrivial aliases correspond to $\ell \neq 0$; independence of coordinates implies the expected phase product vanishes. Orthogonality yields the MSE expression with $1/m$ decay. □

6. BM Regularization: Density Control with a Closed Proof

In discrete settings, invertibility on a chosen band is guaranteed by coefficient floors ($m_K(g) > 0$ or $m_K^{(d)}(G) > 0$). BM regularization targets a complementary goal: maintain *positive BM density of active indices* as the model scales (increasing K and/or infinite Taylor support), aligning with the continuous MIT injectivity theory.

6.1. Hard Band Floors (Guaranteed Frames)

Definition 6.1 (Coefficient-floor parameterization on $\{0,\ldots,K\}$)**.** *Fix $\mu > 0$. Parameterize*

$$c_k = \mu\, e^{i\phi_k}\big(1 + \mathrm{softplus}(a_k)\big), \qquad k = 0,\ldots,K,$$

with trainable real (a_k, ϕ_k).

Then $|c_k| \geq \mu$ and thus $m_K(g) \geq \mu$, yielding guaranteed bounds (6).

6.2. Soft Gap Regularizers (Practical BM Proxies)

Let $g(z) = \sum_{k\geq 0} c_k z^k$ be (possibly infinite) and fix a threshold $\varepsilon > 0$. Define the "active-above-threshold" set

$$S_\varepsilon(g) := \{k \in \mathbb{N} : |c_k| \geq \varepsilon\}.$$

A training objective can penalize long gaps in S_ε.

Definition 6.2 (Hard gap constraint (idealized) and a soft penalty)**.** *Fix $L \in \mathbb{N}$. The* hard *constraint is: $S_\varepsilon(g)$ is L-syndetic. A differentiable surrogate is*

$$\mathcal{R}_{\mathrm{gap}}(g) := \sum_{t=0}^{K_{\max}-L} \max\left(0,\ 1 - \sum_{k=t}^{t+L-1} \sigma\!\left(\frac{|c_k| - \varepsilon}{\tau}\right)\right),$$

where σ is a sigmoid and $\tau > 0$ is a temperature. In the idealized limit $\tau \downarrow 0$, the penalty enforces L-syndeticity of S_ε on $\{0,\ldots,K_{\max}\}$.

6.3. The Key Lemma: Bounded Gaps Imply Positive BM Density

This is the "missing lemma" in earlier drafts; we now prove it fully.

Theorem 6.3 (Syndeticity $\Rightarrow$ positive interior BM density with constants)**.** *Let $S \subset \mathbb{N}$ be L-syndetic. View S as a discrete subset of $\mathbb{R}_{\geq 0}$. Then*

$$D_{BM}(S) \ \geq\ \frac{1}{L}.$$

In particular, any coefficient index set $S(g)$ that contains an L-syndetic subset has positive interior BM density.

Proof. Because S is L-syndetic, for each block $B_m = \{mL,\ldots,(m+1)L-1\}$ choose one element

$$s_m := \min(S \cap B_m).$$

Define the subsequence $\Lambda_0 = \{s_m : m \in \mathbb{N}\} \subset S$. For $x \in [mL,(m+1)L)$ we have

$$m \leq n_{\Lambda_0}(x) \leq m+1, \qquad \text{while} \qquad \frac{x}{L} \in [m, m+1).$$

Hence $|n_{\Lambda_0}(x) - x/L| \leq 1$ for all $x \geq 0$. Therefore

$$\int_0^\infty \frac{|n_{\Lambda_0}(x) - \frac{1}{L}x|}{1+x^2}\,dx \leq \int_0^\infty \frac{1}{1+x^2}\,dx < \infty,$$

 doi:10.20944/preprints202603.1171.v1

so Λ_0 is strongly $(1/L)$-regular in the sense of Definition 2.5. By Definition 2.7, $D_{BM}(S) \geq 1/L$. □

Corollary 6.4 (Gap penalties imply positive BM density (operational statement)). *Fix $\varepsilon > 0$ and suppose training enforces that $S_\varepsilon(g)$ is L-syndetic for some L. Then $S(g)$ has $D_{BM}(S(g)) \geq 1/L > 0$.*

Proof. $S_\varepsilon(g) \subset S(g)$ and Theorem 6.3 applies. □

Lemma 6.5 (Syndeticity implies positive lower counting density). *Let $S \subset \mathbb{N}$ be L-syndetic. Then its (one-sided) lower counting density satisfies*

$$\liminf_{R\to\infty} \frac{\#(S \cap [0,R])}{R} \;\geq\; \frac{1}{L}.$$

Moreover, its upper counting density is finite (indeed $\limsup_{R\to\infty} \#(S \cap [0,R])/R \leq 1$).

Proof. For each $R > 0$, partition $[0,R]$ into $\lfloor R/L \rfloor$ disjoint integer blocks of length L (plus a remainder). Each full block contains at least one element of S, hence

$$\#(S \cap [0,R]) \;\geq\; \left\lfloor \frac{R}{L} \right\rfloor.$$

Divide by R and take $\liminf_{R\to\infty}$ to obtain the bound. The upper density bound is trivial since $\#(S \cap [0,R]) \leq \#(\mathbb{N} \cap [0,R]) \leq R+1$. □

Remark 6.6 (Two density notions, one hierarchy). *For L-syndetic sets $S \subset \mathbb{N}$ we now have* two *explicit lower bounds:*

$$D_{BM}(S) \;\geq\; \frac{1}{L} \qquad \text{(Theorem 6.3)} \quad \text{and} \quad \rho_*(S) \;\geq\; \frac{1}{L} \qquad \text{(Lemma 6.5)}.$$

In our restated MIT theorem (Theorem 7.1), the required density hypothesis is formulated in terms of the counting density ρ_*; the BM bound is a stronger *regularity statement that comes "for free" from bounded gaps in this model.*

6.4. A Bounded-Gap Equivalence: Strong Regularity Versus Weighted Lattice Deviations

The definition of strong a-regularity (2) is "continuum" (an L^1-type integral of counting-function deviations). For bounded-gap sequences one can convert it to an equivalent *discrete* weighted summability statement, which is convenient both conceptually and computationally.

Setup

Let $\Lambda \subset \mathbb{R}_{\geq 0}$ be discrete and enumerate it increasingly as

$$0 \leq \lambda_1 < \lambda_2 < \cdots, \qquad \lambda_n \to \infty.$$

Assume Λ has bounded gaps: there exists $L > 0$ such that

$$\lambda_{n+1} - \lambda_n \leq L \qquad \text{for all } n \geq 1. \tag{24}$$

For $a > 0$ define the lattice deviations

$$\Delta_n^{(a)} := \lambda_n - \frac{n}{a}.$$

Lemma 6.7 (Strong regularity forces asymptotic slope). *If Λ is strongly a-regular for some $a > 0$, then $\lim_{x\to\infty} \frac{n_\Lambda(x)}{x} = a$ and hence $\lambda_n \sim n/a$.*

Proof. Suppose $\limsup_{x\to\infty} \frac{|n_\Lambda(x)-ax|}{x} > 0$. Then there exist $\varepsilon > 0$ and a sequence $R_j \to \infty$ such that $|n_\Lambda(R_j) - aR_j| \geq \varepsilon R_j$. Choosing a subsequence so that the intervals $[R_j, 2R_j]$ are disjoint, we obtain

$$\int_{R_j}^{2R_j} \frac{|n_\Lambda(x) - ax|}{1+x^2}\, dx \;\geq\; \varepsilon R_j \int_{R_j}^{2R_j} \frac{dx}{1+x^2} \;\gtrsim\; \varepsilon,$$

and summing over j forces the integral in (2) to diverge, a contradiction. Thus $|n_\Lambda(x) - ax| = o(x)$, i.e. $n_\Lambda(x)/x \to a$. Since $n_\Lambda(\lambda_n) = n$, we have $n/(a\lambda_n) \to 1$, i.e. $\lambda_n \sim n/a$. □

Theorem 6.8 (Series characterization of strong a-regularity under bounded gaps)**.** *Let Λ satisfy* (24) *and fix $a > 0$. Then Λ is strongly a-regular (Definition 2.5) if and only if*

$$\sum_{n\geq 1} \frac{|\Delta_n^{(a)}|}{1+n^2} < \infty. \tag{25}$$

Proof. Write $n(x) = n_\Lambda(x)$. For $x \in [\lambda_n, \lambda_{n+1})$ one has $n(x) = n$, hence

$$\int_0^\infty \frac{|n(x) - ax|}{1+x^2}\, dx = \sum_{n\geq 1} \int_{\lambda_n}^{\lambda_{n+1}} \frac{|n-ax|}{1+x^2}\, dx. \tag{26}$$

Step 1: local weight comparability on each interval.

Since $x \in [\lambda_n, \lambda_{n+1}]$ and $\lambda_{n+1} - \lambda_n \leq L$, we have

$$\frac{1}{1+x^2} \asymp \frac{1}{1+\lambda_n^2} \qquad \text{uniformly for } x \in [\lambda_n, \lambda_{n+1}],$$

with constants depending only on L. (The finitely many small n contribute only a finite constant.)

Step 2: isolate the deviation $\Delta_n^{(a)}$.

For $x = \lambda_n + t$ with $t \in [0, \lambda_{n+1} - \lambda_n]$,

$$n - ax = a\Big(\frac{n}{a} - \lambda_n\Big) - at = -a(\Delta_n^{(a)} + t).$$

Thus, using Step 1 and $\lambda_{n+1} - \lambda_n \leq L$,

$$\begin{aligned}\int_{\lambda_n}^{\lambda_{n+1}} \frac{|n-ax|}{1+x^2}\, dx &\asymp \frac{1}{1+\lambda_n^2} \int_0^{\lambda_{n+1}-\lambda_n} |\Delta_n^{(a)} + t|\, dt \\ &\asymp \frac{|\Delta_n^{(a)}|}{1+\lambda_n^2} + \frac{1}{1+\lambda_n^2},\end{aligned} \tag{27}$$

where the implicit constants depend only on a and L.

Step 3: ($\Rightarrow$) strong regularity implies the series.

Assume Λ is strongly a-regular. Then Lemma 6.7 gives $\lambda_n \sim n/a$, so $\sum_n (1+\lambda_n^2)^{-1} < \infty$ and $(1+\lambda_n^2)^{-1} \asymp (1+n^2)^{-1}$ for large n. Summing (27) over n and using (26), we obtain $\sum_n \frac{|\Delta_n^{(a)}|}{1+\lambda_n^2} < \infty$, hence (25).

Step 4: ($\Leftarrow$) the series implies strong regularity.

Assume (25). Then $|\Delta_n^{(a)}| = o(n)$, so $\lambda_n = n/a + o(n)$ and again $(1+\lambda_n^2)^{-1} \asymp (1+n^2)^{-1}$ and $\sum_n (1+\lambda_n^2)^{-1} < \infty$. Therefore $\sum_n \frac{|\Delta_n^{(a)}|}{1+\lambda_n^2} < \infty$. Summing (27) and using (26) gives $\int_0^\infty \frac{|n(x)-ax|}{1+x^2}\, dx < \infty$, i.e. Λ is strongly a-regular. □

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1171.v1

Remark 6.9 (Computable surrogates for strong regularity). *For integer index sets $S \subset \mathbb{N}$ with bounded gaps, Theorem 6.8 turns strong regularity into a weighted summability of deviations from an arithmetic progression. This motivates practical "density control" penalties such as*

$$\mathcal{R}_a(S;N) := \sum_{n=1}^{N} \frac{|s_n - n/a|}{1+n^2},$$

where (s_n) enumerates S and a is a target activity rate. Unlike a hard floor constraint on coefficients, this penalty can be optimized smoothly (after replacing $|\cdot|$ by a differentiable proxy) and encourages a near-lattice structure of active indices. Related discrete criteria for BM density and spectral gaps appear in [3].

7. Continuum Bridge: From Gap Regularization to MIT Injectivity

We now close the continuum bridge: a practical gap constraint on the learned entire generator implies positive interior Beurling–Malliavin density (a strong structural regularity) and, in particular, the positive lower counting density $\rho_* > 0$ required by the MIT injectivity theorem stated below.

7.1. The MIT Orbit-Completeness and Injectivity Theorem (External Input)

We now restate the precise theorem we use from [1], with notation aligned to that paper.

Theorem 7.1 (MIT orbit completeness and global injectivity (cf. [1], Thm. 2–3)). *Let $-1 < \Re p < 1$ and let f be p-admissible in the sense of [1]. Let g be an entire function of finite order, expanded about α as*

$$g(\alpha+z) = \sum_{k=0}^{\infty} c_k z^k, \qquad S := \{k \in \mathbb{N} : \ c_k \neq 0\}.$$

Assume the density and boundedness hypotheses of [1, Assumptions 2–3]:

$$0 < \rho_* := \liminf_{R\to\infty} \frac{\#(S\cap[0,R])}{R} \le \rho^* := \limsup_{R\to\infty} \frac{\#(S\cap[0,R])}{R} < \infty,$$

and g is bounded on an annulus $r_- < |z-\alpha| < r_+$ with $0 < r_- < |\beta| < r_+$, and the (harmless) normalization $c_0 = 0$, $c_1 \neq 0$ holds. Define the Master Integral Transform (MIT) by

$$G^p_{\alpha,\beta}[f](\theta) := \frac{1}{2\pi} \int_{\mathbb{R}} |x|^p f(x) e^{-i\pi p \operatorname{sgn}(x)/2} g(\alpha + \beta e^{i\theta x})\, dx.$$

Then the orbit $\Theta_g := \{x \mapsto g(\alpha+\beta e^{i\theta x}) : \theta \in \mathbb{R}\}$ is complete in $L^2_*(\mathbb{R})$, and in particular

$$G^p_{\alpha,\beta}[f](\theta) \equiv 0 \ \ \forall \theta \in \mathbb{R} \quad \Longrightarrow \quad f = 0 \text{ a.e. on } \mathbb{R}.$$

Remark 7.2. *The theorem is stated here because it is the precise analytic "target" that our gap-regularization mechanism aims to satisfy: the only nontrivial input is the positivity of the lower density ρ_* of the nonzero Taylor indices.*

7.2. Unconditional Theorem: Gap Regularization Implies MIT Injectivity

Theorem 7.3 (Gap regularization $\Rightarrow$ MIT injectivity). *Let g be entire of finite order and fix $\varepsilon > 0$. Suppose training enforces that $S_\varepsilon(g)$ is L-syndetic for some L (e.g. via a gap penalty or a hard bounded-gap constraint). Then:*

1. *$S(g)$ has positive interior BM density (indeed $D_{BM}(S(g)) \ge 1/L$).*
2. *In particular, the lower density ρ_* of [1] satisfies $\rho_* \ge 1/L > 0$.*

Consequently, provided the remaining hypotheses of [1] hold (boundedness of g on an annulus around α with $0 < r_- < |\beta| < r_+$ and the harmless normalization $c_0 = 0$, $c_1 \neq 0$), the associated MIT is complete and globally injective in the sense of Theorem 7.1, and admits the explicit inversion formulae of [1, Section 7].

Proof. Item (1) is Corollary 6.4. For item (2), if $S_\varepsilon(g)$ is L-syndetic then every interval of integers of length L contains at least one element of $S_\varepsilon(g)$; hence for all R,

$$\#(S(g) \cap [0, R]) \ \geq \ \#(S_\varepsilon(g) \cap [0, R]) \ \geq \ \left\lfloor \frac{R}{L} \right\rfloor,$$

which implies $\rho_* \geq 1/L$. Apply Theorem 7.1. □

Remark 7.4 (Interpretation for AI). *This theorem provides the end-to-end conceptual guarantee: a* computable *gap constraint on learned Taylor coefficients enforces the* analytic density hypothesis *behind MIT injectivity. In discrete implementations, coefficient floors guarantee invertibility on finite bands; in continuum settings, syndetic coefficient activity pushes the kernel into the MIT-injective regime.*

8. Beyond-Density Injectivity Mechanisms for the MIT

The MIT injectivity theorem (Theorem 7.1) gives a sufficient density hypothesis on the set of nonzero Taylor indices of the kernel generator. As emphasized in [1], this hypothesis is not necessary: classical Fourier/Laplace kernels fall outside the density regime yet remain injective by classical harmonic analysis. This section records two complementary injectivity mechanisms that *do not* require asymptotic density.

8.1. A β-Analytic Injectivity Theorem (Density Not Needed)

Recall the weighted signal

$$h_p(x) := |x|^p f(x)\, e^{-i\frac{\pi p}{2}\operatorname{sgn}(x)}.$$

We use the Fourier transform convention

$$\widehat{h}(\omega) := \int_{\mathbb{R}} h(x)\, e^{-i\omega x}\, dx, \qquad h(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{h}(\omega)\, e^{i\omega x}\, d\omega \ \ (\text{when } h, \widehat{h} \in L^1).$$

For an entire g and parameters $\alpha, \beta \in \mathbb{C}$, the MIT is

$$G^p_{\alpha,\beta}[f](\theta) := \frac{1}{2\pi}\int_{\mathbb{R}} h_p(x)\, g(\alpha + \beta e^{i\theta x})\, dx, \tag{28}$$

whenever the integral converges.

Theorem 8.1 (β-analytic injectivity for any nonconstant entire kernel)**.** *Let g be a nonconstant entire function and fix $\alpha \in \mathbb{C}$. Assume $h_p \in L^1(\mathbb{R})$. Suppose that for each fixed $\theta \in \mathbb{R}$, the function $\beta \mapsto G^p_{\alpha,\beta}[f](\theta)$ is known for all β in some neighborhood of* 0 *(as an analytic function of β). Then f is uniquely determined by this data.*

More explicitly, for any $k \geq 1$,

$$\partial^k_\beta\, G^p_{\alpha,\beta}[f](\theta)\big|_{\beta=0} = \frac{g^{(k)}(\alpha)}{2\pi}\int_{\mathbb{R}} h_p(x)\, e^{ik\theta x}\, dx = \frac{g^{(k)}(\alpha)}{2\pi}\, \widehat{h_p}(-k\theta), \tag{29}$$

so if $g^{(k_0)}(\alpha) \neq 0$ for some $k_0 \geq 1$, then $\widehat{h_p}$ is recovered by the change of variables $\omega = -k_0\theta$.

Proof. Since g is entire, for each fixed θ the map $\beta \mapsto g(\alpha + \beta e^{i\theta x})$ is analytic and admits the Taylor expansion

$$g(\alpha + \beta e^{i\theta x}) = \sum_{k=0}^{\infty} \frac{g^{(k)}(\alpha)}{k!}\, \beta^k e^{ik\theta x},$$

which converges uniformly in x on compact β-discs. If $|\beta|$ is small enough then $\alpha + \beta e^{i\theta x}$ ranges in a compact set independent of x (because $e^{i\theta x}$ lies on the unit circle), hence $|g(\alpha + \beta e^{i\theta x})| \leq C$ for some constant C. Because $h_p \in L^1(\mathbb{R})$, dominated convergence justifies differentiating under the integral sign in (28) at $\beta = 0$, yielding (29). If g is nonconstant then not all derivatives $g^{(k)}(\alpha)$ vanish; choose

$k_0 \geq 1$ with $g^{(k_0)}(\alpha) \neq 0$. Knowing $\partial_\beta^{k_0} G^p_{\alpha,\beta}[f](\theta)|_{\beta=0}$ for all θ gives $\widehat{h_p}(\omega)$ for all $\omega \in \mathbb{R}$, hence h_p (and f) by Fourier inversion. □

Remark 8.2 (Interpretation)**.** *Theorem 8.1 shows that once the experiment provides multi-β information (several β-channels near 0, or the ability to form finite differences), injectivity becomes a purely analytic property of g—and any nonconstant entire generator suffices. The hard case addressed by Theorem 7.1 is the one-parameter regime with fixed β.*

8.2. Finite-Band Injectivity from Generic Shifts (no Density on Windows)

Even with fixed β, invertibility on a finite band only requires the first $K+1$ Taylor coefficients about the chosen center to be nonzero. For a nonpolynomial entire generator, this is generically achieved by shifting the expansion center.

Theorem 8.3 (Generic shift yields full Taylor support on any finite band)**.** *Let g be a nonpolynomial entire function and fix $K \in \mathbb{N}$. Define the exceptional set*

$$E_K := \bigcup_{k=0}^{K} \{\alpha \in \mathbb{C} : g^{(k)}(\alpha) = 0\}.$$

Then E_K is a finite union of closed discrete sets, hence has empty interior. In particular, $\mathbb{C} \setminus E_K$ is open and dense, and for any $\alpha \in \mathbb{C} \setminus E_K$ the Taylor coefficients of $g(\alpha + z)$ satisfy

$$c_k(\alpha) = \frac{g^{(k)}(\alpha)}{k!} \neq 0 \qquad \text{for } k = 0, 1, \ldots, K.$$

Consequently, replacing $g(z)$ by the shifted generator $g_\alpha(z) = g(\alpha + z)$, the discrete oversampled MIT layer on $\mathrm{PW}_K(\mathbb{T})$ is exactly invertible with the same frame bounds and conditioning statements as in Theorems 3.3 and 3.5.

Proof. Because g is nonpolynomial, each derivative $g^{(k)}$ is not identically zero. Zeros of a nontrivial entire function form a closed discrete subset of $\mathbb{C}$ (no accumulation point in $\mathbb{C}$). A finite union of closed discrete sets is closed and has empty interior. Choosing $\alpha \notin E_K$ yields $g^{(k)}(\alpha) \neq 0$ for $k \leq K$, hence $c_k(\alpha) \neq 0$ and the discrete frame theory applies verbatim. □

Remark 8.4. *Theorem 8.3 is a practical "escape hatch": even if the Taylor support at a default center is sparse, shifting α generically restores full support on any finite band. In ML parameterizations, α can be learned alongside the Taylor coefficients.*

9. Entire-Generated Nonharmonic Systems and a Monomial Rigidity Principle

This section isolates a structural fact: except for the monomial case, entire generators produce *nonharmonic* measurement atoms and orbit systems (each atom contains multiple Fourier modes). We also characterize the Gram-design cone of these atoms on finite bands.

9.1. Frame Elements and the Gram Design Cone

Fix $K < M$ and define the band-projected kernel functions (frame elements)

$$\varphi_r(x) := \sum_{k=0}^{K} c_k \, \beta_r^k \, e^{-ikx}, \qquad r = 0, 1, \ldots, M-1, \tag{30}$$

which are precisely the PW_K projections of $g(\beta_r e^{-ix})$.

Proposition 9.1 (Frame operator is diagonal in Fourier coefficients)**.** *For $f \in \mathrm{PW}_K(\mathbb{T})$,*

$$(\mathcal{A}^{(M)}_{g,K} f)_r = \langle f, \varphi_r \rangle_{L^2(\mathbb{T})}.$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1171.v1

Moreover, $\{\varphi_r\}_{r=0}^{M-1}$ *is a frame for* $\mathrm{PW}_K(\mathbb{T})$ *whose frame operator* S *acts diagonally on one-sided Fourier coefficients:*

$$\widehat{Sf}(k) = M|c_k|^2 \hat{f}(k), \qquad 0 \le k \le K.$$

Proof. The inner-product identity follows by expanding φ_r and using orthogonality of e^{ikx} on $\mathbb{T}$. In coefficient coordinates, $\mathcal{A}_{g,K}^{(M)} = V_{M,K} D_g$ (Section 3), hence

$$S = (VD_g)^*(VD_g) = D_g^*(V^*V)D_g = D_g^*(MI)D_g,$$

which yields the diagonal action and the displayed formula. □

Theorem 9.2 (Circulant Gram design cone). *Let* $G \in \mathbb{C}^{M\times M}$ *be the Gram matrix* $G_{rs} = \langle \varphi_r, \varphi_s \rangle$*. Then* G *is circulant and diagonalized by the* $M \times M$ *DFT matrix* F_M*:*

$$G = F_M^* \operatorname{diag}(\lambda)\, F_M, \qquad \lambda_k = M|c_k|^2 \text{ for } 0 \le k \le K, \text{ and } \lambda_k = 0 \text{ for } k > K.$$

Conversely, for any nonnegative vector $\lambda \in \mathbb{R}_{\ge 0}^M$ *supported on* $\{0,\dots,K\}$*, there exists a choice of coefficients* $\{c_k\}_{k=0}^K$ *producing exactly this Gram spectrum. Equivalently, the set of Gram matrices realizable by entire-generated* PW_K *frames with roots-of-unity sampling is exactly the cone of positive semidefinite circulant matrices whose DFT-spectrum is supported in* $\{0,\dots,K\}$*.*

Proof. Using (30),

$$\langle \varphi_r, \varphi_s \rangle = \sum_{k=0}^{K} |c_k|^2 \beta_r^k \overline{\beta_s^k} = \sum_{k=0}^{K} |c_k|^2 \omega_M^{(r-s)k},$$

which depends only on $r - s$ modulo M, hence G is circulant. DFT diagonalization of circulants gives eigenvalues λ_k, and a direct computation yields $\lambda_k = M|c_k|^2$ for $k \le K$ and 0 for $k > K$. Conversely, given such λ, set $|c_k|^2 = \lambda_k/M$ for $k \le K$ and choose arbitrary phases. □

9.2. A Monomial Rigidity Principle

The orbit functions $x \mapsto g(\beta e^{-ix})$ reduce to *single-frequency* Fourier exponentials only in the monomial case.

Theorem 9.3 (Monomial rigidity). *Let* $g(z) = \sum_{k=0}^{\infty} c_k z^k$ *be entire and fix any* $\beta \neq 0$*. If the function* $x \mapsto g(\beta e^{-ix})$ *is a scalar multiple of a single character* e^{-inx}*, then all but one coefficient* c_k *vanish; i.e.* $g(z) = c_{k_0} z^{k_0}$ *is a monomial. Equivalently: any entire generator with at least two nonzero Taylor coefficients produces a genuinely nonharmonic orbit element.*

Proof. Write

$$g(\beta e^{-ix}) = \sum_{k=0}^{\infty} c_k \beta^k e^{-ikx}.$$

If this equals Ce^{-inx} for some $C \in \mathbb{C}$ and integer n, then the Fourier series of the left-hand side has support contained in $\{n\}$. By uniqueness of Fourier coefficients on $\mathbb{T}$, we obtain $c_k\beta^k = 0$ for all $k \neq n$, hence $c_k = 0$ for all $k \neq n$. □

Remark 9.4 (Relation to Fourier-multiplier layers). *A Fourier multiplier* T *satisfies* $T(e^{-inx}) = m(n)e^{-inx}$*, so it maps characters to (scaled) characters. By Theorem 9.3, an entire-generated atom is a character only in the monomial case, hence cannot be obtained by applying a translation-invariant multiplier to a single character in the nonmonomial regime. This does not say that Fourier-based architectures cannot represent the same* linear span *(they can); rather, it highlights that entire generators provide a structured way to produce multi-harmonic atoms from a shared analytic parameterization.*

10. AI Applications and Implementation Notes

This section translates the analytic operator $\mathcal{A}_{G,K}^{(d,M)}$ into a practical module for machine learning and signal processing. The guiding idea is simple but unusually useful:

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1171.v1

Choose an entire generator G whose Fourier coefficients $\{c_\alpha\}$ are nonvanishing on the target band $[0..K]^d$ (and, in practice, are bounded away from 0 by a floor). Then $\mathcal{A}_{G,K}^{(d,M)}$ is a structured, FFT-like, multi-channel feature map that is (i) exactly invertible *on the K-band and (ii) admits* provably unbiased *Monte–Carlo de-aliasing when the input has energy beyond the band.*

10.1. Where the "Learning" Lives

In a neural pipeline, there are three natural places to introduce trainable parameters while keeping the analysis–reconstruction guarantees:

1. **Train the generator coefficients** $\{c_\alpha\}$ (or a low-dimensional parametrization of them) subject to the non-vanishing / floor constraint on $[0..K]^d$.
2. **Train a downstream classifier** on top of either (i) the measurement vector $y = \mathcal{A}_{G,K}^{(d,M)} f$ itself, or (ii) the recovered band coefficients $\widehat{\widetilde{f}}(\alpha)$ produced by the inverse formula in Theorem 5.3.
3. **Train the band limit** K and the channel count M (or run multi-resolution), using the de-aliasing theory to maintain stability.

The rest of the paper provides the analytic "guard rails": band invertibility, quantitative stability via the floors, and a variance-controlled de-aliasing mechanism that turns the inevitable non-bandlimited reality of data into a manageable noise term.

10.2. Image Classification: Compressed Spectral Measurements on $\mathbb{T}^2$

For an image $f : \mathbb{T}^2 \to \mathbb{R}$, the layer outputs

$$y = \mathcal{A}_{G,K}^{(2,M)} f \in \mathbb{C}^{M^2},$$

with an inversion (on the K-band) given by

$$\widehat{f}(\alpha) \approx \frac{1}{c_\alpha} \frac{1}{M^2} \sum_{r \in [0..M-1]^2} y_r \, \beta_r^{-\alpha}, \qquad \alpha \in [0..K]^2.$$

When the input is not strictly K-bandlimited (which is the default for real images), the residue-class mixing described in Lemma 4.1 appears, and the Monte–Carlo rotation trick from Theorem 5.5 can be used to suppress it.

A concrete end-to-end supervised pipeline is:

1. Choose (K, M) and a generator G (fixed or learnable-with-floors).
2. Compute $y = \mathcal{A}_{G,K}^{(2,M)} f$ (or multiple rotated copies $y^{(j)}$).
3. Recover a stable band representation $\widehat{\widetilde{f}}(\alpha)$ using (optionally) the de-aliasing estimator.
4. Feed $\{\Re \widehat{\widetilde{f}}(\alpha), \Im \widehat{\widetilde{f}}(\alpha)\}_{\alpha \in [0..K]^2}$ to a classifier.

Section 11 gives a full-data example on handwritten digits with a deliberately adversarial high-frequency corruption. Figures 1 and 2 and Table 2 summarize the outcome: naive aliasing degrades robustness, while de-aliasing (increasing m) recovers the Fourier-band signal and restores accuracy.

10.3. ECG/EEG-Style 1D Signals: Robust Band Recovery and Classification

For a 1D signal $f : \mathbb{T} \to \mathbb{R}$ (ECG, EEG, EMG, vibration, audio, ...), the measurement

$$y = \mathcal{A}_{g,K}^{(1,M)} f \in \mathbb{C}^M$$

can be used either as a feature vector directly, or inverted (on the band) to obtain stable Fourier-band coefficients. In biomedical settings, the relevant nuisance is often *out-of-band corruption* (sensor noise, baseline wander, muscle artifacts, power-line harmonics). The de-aliasing estimator turns such corruption into a controlled variance term.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1171.v1

In Section 11, we provide a full example on an ECG record (windowed classification of R-peak alignment) under additive white/pink/brown noise at 0 dB SNR, a regime in which many simplistic representations break. Figure 4 shows representative windows, while Figure 5 and Table 3 quantify the benefit of increasing m.

10.4. Implementation: A De-Aliasing "Wrapper" for any MIT Front-End

Theorem 5.5 provides an estimator with $\mathcal{O}(1/m)$ mean-square error decay. Algorithm 1 gives a practical recipe.

Algorithm 1 MIT-DeAlias: Monte–Carlo de-aliasing for band recovery

Input: band limit K, channel count $M \geq K+1$, generator coefficients $\{c_k\}$ (or c_α), number of rotations m
Input: a signal/image f (or a batch) and the ability to evaluate $\mathcal{A}^{(d,M)}_{G,K}$
Sample i.i.d. random rotations $a^{(j)} \in \mathbb{T}^d$, $j = 1, \dots, m$
for $j = 1$ to m **do**
    Compute measurements $y^{(j)} = \mathcal{A}^{(d,M)}_{G,K} f$ using $\beta^{(j)}_r = a^{(j)} \odot \beta_r$
    Recover band coefficients $\widetilde{\widehat{f}}^{(j)}(\alpha)$ by inverting the K-band system (Theorem 5.3)
end for
Output the averaged estimate $\widetilde{\widehat{f}}^{\mathrm{MC}}(\alpha) = \frac{1}{m}\sum_{j=1}^{m} \widetilde{\widehat{f}}^{(j)}(\alpha)$

In practice one can use the same idea for mini-batches, amortize the Fourier matrices, and (when learning G) enforce floors by parametrization (e.g. $c_k = \mu + \mathrm{softplus}(\theta_k)$). Figure 7 illustrates why floors matter when measurements are noisy: coefficients near zero amplify inversion noise and can substantially reduce downstream accuracy.

11. Numerical Experiments

This section presents full-data experiments on (i) images and (ii) biomedical-style time series. All experiments follow the same principle: we train on *clean* data and evaluate on *corrupted* data, so the reported numbers quantify robustness rather than mere interpolation.

11.1. General Experimental Protocol

Data splitting and metrics.

We use stratified 5-fold cross-validation. For multi-class tasks we report mean ± standard deviation of *accuracy* and *macro-F1*. For binary tasks we report mean ± standard deviation of *accuracy* and *F1*, and also the mean AUC (area under the ROC curve).

Classifier.

To focus on the representation (not on a heavy downstream model), we use a linear SVM (standardized features) in all experiments.

MIT parameters.

Unless stated otherwise, we use band limit K and channels $M \geq K+1$ as indicated below. Generators are of the form

$$c_n = \frac{\tau^n}{(n!)^s} \qquad (s > 0)$$

in 1D and

$$c_\alpha = \frac{\tau^{|\alpha|}}{\prod_{j=1}^{d}(\alpha_j!)^s}$$

 doi:10.20944/preprints202603.1171.v1

in d dimensions, which defines an entire function for every $s > 0$. This family conveniently interpolates between rapidly decaying exponential-type coefficients ($s = 1$) and heavier-tailed entire generators ($0 < s < 1$).

11.2. Experiment I: Image Classification Under High-Frequency Corruption (Digits)

Dataset.

We use the standard handwritten digits dataset distributed with `scikit-learn` (1797 images, 10 classes, 8×8 grayscale). [14]

Corruption model.

Each test image is corrupted by adding a fixed high-frequency checkerboard pattern (a pure $(4, 4)$ Fourier mode on the 8×8 grid) with amplitude $\varepsilon = 6$. This perturbation is intentionally chosen so that, when using a coarse channel count $M = 4$, it aliases directly into the low band $[0..3]^2$.

MIT setup.

We take $K = 3$, $M = 4$ and the entire generator family with $(\tau, s) = (2, 1)$ (exponential-type decay). We compare the naive estimator ($m = 1$) and the Monte–Carlo estimator with increasing m.

Results.

Figure 1 shows representative clean/corrupted samples. Table 2 reports robust classification performance (train on clean, test on corrupted). Figure 2 shows that increasing m restores robustness, approaching the FFT low-band baseline, while wavelet and raw-pixel baselines degrade sharply under this adversarial high-frequency perturbation. Finally, Figure 3 empirically confirms the $\mathcal{O}(1/m)$ decay of the band-recovery MSE predicted by Theorem 5.5.

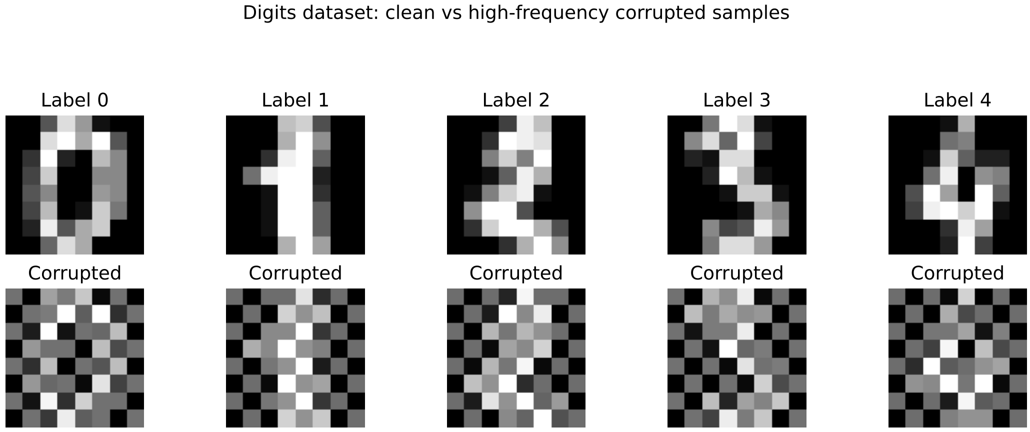

Figure 1. Digits dataset (images from `scikit-learn`): clean vs deliberately high-frequency corrupted samples. The corruption is designed to alias into the coarse-band recovery when $M = 4$.

Table 2. Digits robustness experiment (train clean, test corrupted). Mean $\pm$ std over 5 folds. MIT de-aliasing (large m) approaches the FFT low-band baseline, while wavelet and raw-pixel representations fail under the adversarial high-frequency corruption.

Method	Dim	Accuracy	MacroF1
Raw pixels (64 dims)	64	0.131 ± 0.040	0.043 ± 0.033
Wavelet db2 (level=2)	169	0.283 ± 0.152	0.233 ± 0.168
FFT low-band (K=3)	32	0.951 ± 0.005	0.951 ± 0.005
MIT naive (M=4,K=3,m=1)	32	0.933 ± 0.013	0.932 ± 0.014
MIT MC (M=4,K=3,m=128)	32	0.950 ± 0.009	0.950 ± 0.009

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2026 doi:10.20944/preprints202603.1171.v1

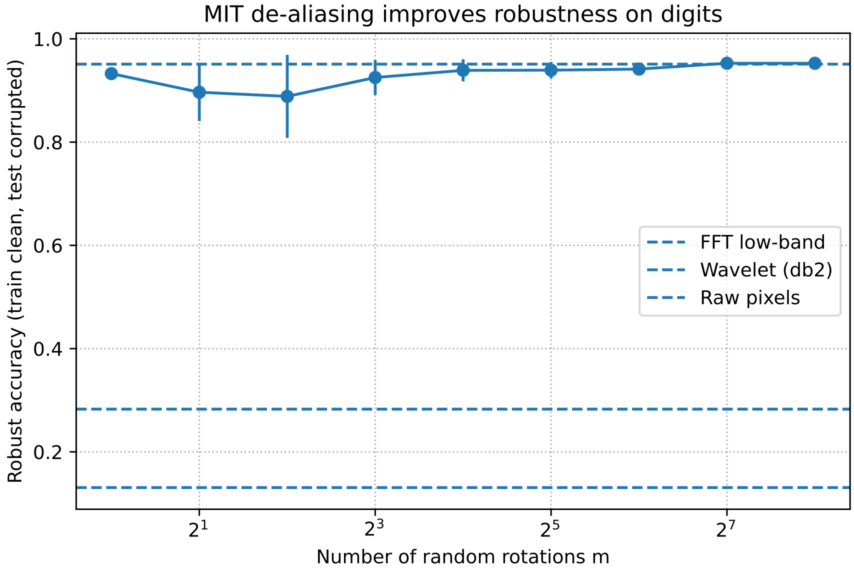

Figure 2. Digits robustness vs number of random rotations m. Error bars reflect variation over different random rotation seeds.

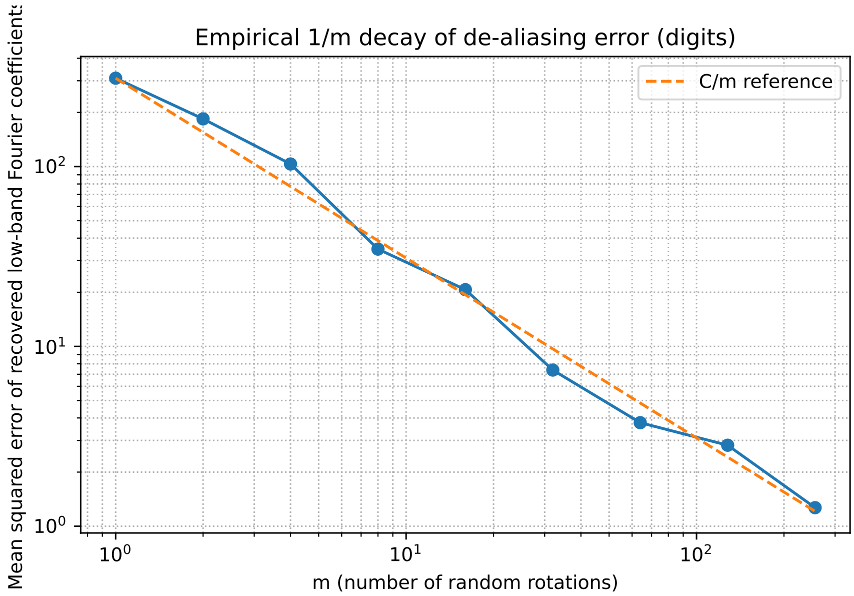

Figure 3. Digits: MSE of recovered low-band Fourier coefficients vs m (log–log). The dashed line is a C/m reference.

11.3. Experiment II: ECG Window Classification Under Colored Noise

Dataset.

We use the ECG record distributed with `scipy.misc.electrocardiogram` (sampling rate 360 Hz; excerpted from MIT-BIH arrhythmia database record 208). [15,16] We apply a standard Butterworth band-pass filter (0.5–40 Hz) and normalize to zero mean and unit variance.

Task: R-centered window classification.

We segment the signal into windows of length $N = 512$ with hop size 256. Using peak detection on the clean filtered signal, we label a window as positive if it contains an R-peak within ± 80 ms of the window center (a simple but practically relevant localization surrogate).

 doi:10.20944/preprints202603.1171.v1

Corruptions.

We create three test sets by adding (independently, per window) white, pink ($1/f$), and brown ($1/f^2$) noise at 0 dB SNR.

MIT setup.

We use $K = 31$, $M = 32$ and an entire generator with parameters $(\tau, s) = (4.5, 0.5)$, which yields moderately heavy tails and thus a nontrivial aliasing regime. We compare naive recovery ($m = 1$) to Monte–Carlo de-aliasing ($m = 64$), and we also report wavelet and FFT low-band baselines.

Results.

Figure 4 shows representative clean/noisy windows. Figure 5 shows that increasing m improves robustness under severe noise. Table 3 summarizes the full results across noise types. Finally, Figure 6 empirically exhibits a near-$1/m$ decay of the band-recovery MSE, consistent with Theorem 5.5.

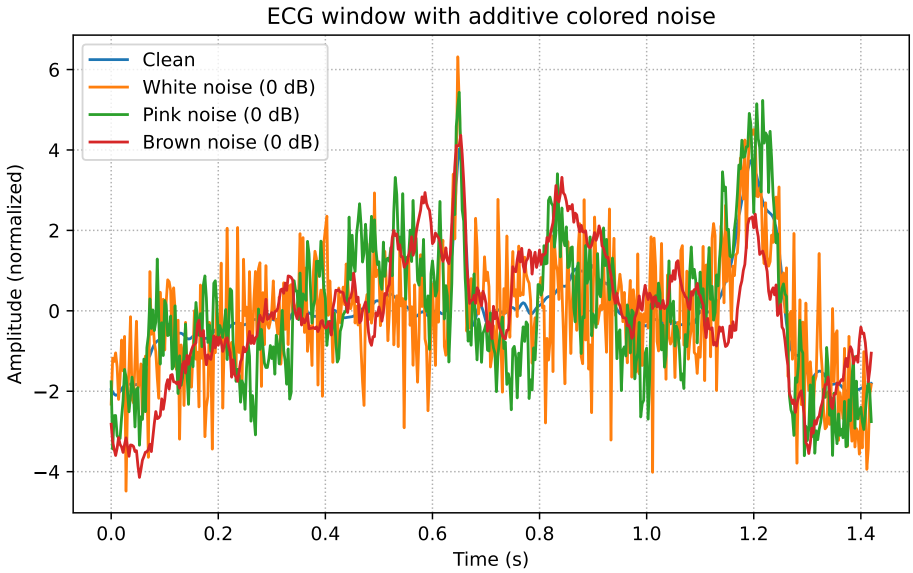

Figure 4. ECG window classification dataset: a representative clean window and its colored-noise corruptions at 0 dB SNR (white/pink/brown).

Table 3. ECG robustness experiment (train clean, test noisy). Mean ± std over 5 folds. MIT Monte–Carlo de-aliasing ($m = 64$) improves robustness over the naive ($m = 1$) estimate and substantially outperforms a wavelet baseline.

Noise	Method	Dim	Acc	F1	AUC
White	Raw time (512 dims)	512	0.836 ± 0.033	0.694 ± 0.062	0.877
White	Wavelet db4 (level=4)	538	0.507 ± 0.044	0.318 ± 0.027	0.502
White	FFT low-band (K=31)	64	0.812 ± 0.039	0.647 ± 0.081	0.896
White	MIT naive (M=32,K=31,m=1)	64	0.752 ± 0.036	0.514 ± 0.087	0.750
White	MIT MC (M=32,K=31,m=64)	64	0.829 ± 0.040	0.658 ± 0.095	0.863
Pink	Raw time (512 dims)	512	0.771 ± 0.050	0.599 ± 0.089	0.813
Pink	Wavelet db4 (level=4)	538	0.533 ± 0.030	0.386 ± 0.039	0.572
Pink	FFT low-band (K=31)	64	0.783 ± 0.041	0.589 ± 0.102	0.830
Pink	MIT naive (M=32,K=31,m=1)	64	0.707 ± 0.052	0.462 ± 0.070	0.684
Pink	MIT MC (M=32,K=31,m=64)	64	0.774 ± 0.037	0.584 ± 0.074	0.815
Brown	Raw time (512 dims)	512	0.757 ± 0.029	0.577 ± 0.055	0.813
Brown	Wavelet db4 (level=4)	538	0.593 ± 0.031	0.420 ± 0.053	0.620
Brown	FFT low-band (K=31)	64	0.788 ± 0.025	0.630 ± 0.042	0.827
Brown	MIT naive (M=32,K=31,m=1)	64	0.750 ± 0.043	0.518 ± 0.061	0.744
Brown	MIT MC (M=32,K=31,m=64)	64	0.783 ± 0.040	0.610 ± 0.065	0.851

 doi:10.20944/preprints202603.1171.v1

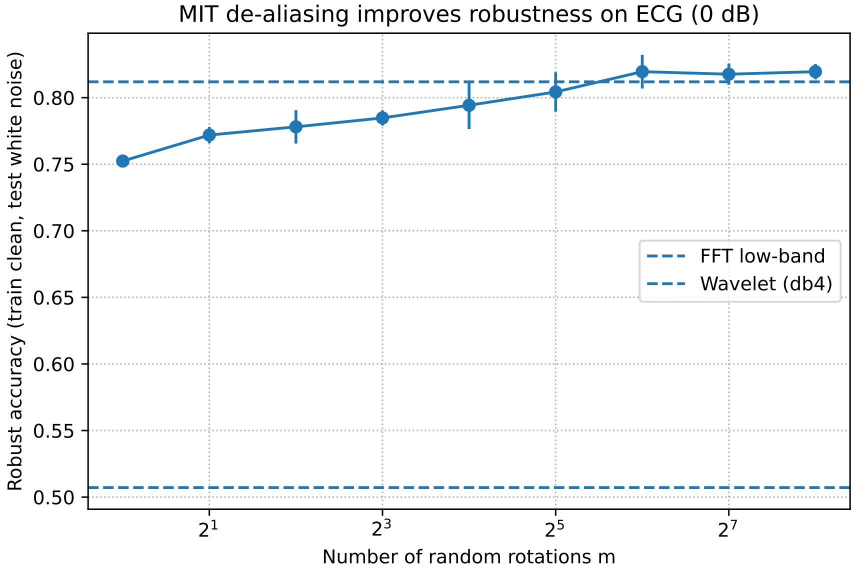

Figure 5. ECG robustness (white-noise corruption): accuracy vs number of random rotations m. Error bars reflect variation over random rotation seeds.

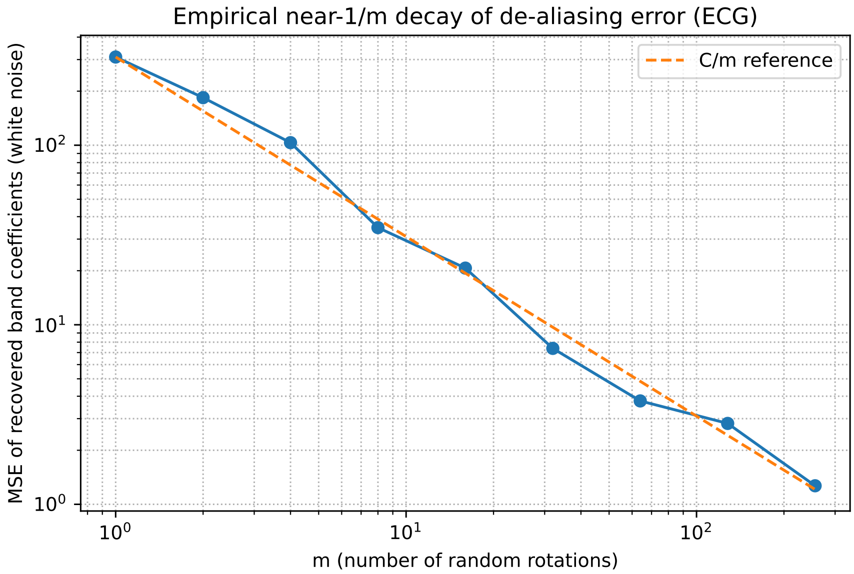

Figure 6. ECG: MSE of recovered band coefficients vs m (log–log) with a C/m reference.

11.4. Stability of Inversion Under Measurement Noise: Why Floors Matter

To isolate the conditioning mechanism described in Section 6, we add small complex Gaussian noise to the measurements before inversion. If some $|c_k|$ are extremely small, the inverse map divides by c_k and amplifies this perturbation, leading to substantial degradation in the recovered band and in downstream classification. Figure 7 illustrates this effect on the ECG task: increasing a floor parameter μ improves robustness.

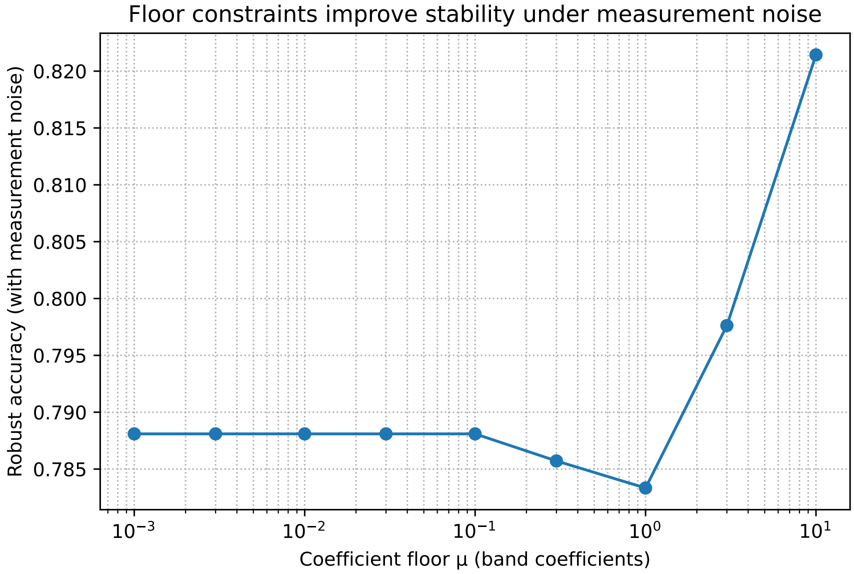

Figure 7. Effect of a coefficient floor μ (applied to the band coefficients) on ECG robustness when measurements are noisy. Floors improve conditioning and downstream accuracy.

12. Conclusion and Open Problems

We built an comprehensive, proof-complete discrete theory for entire-kernel spectral layers on $\mathbb{T}$ and $\mathbb{T}^d$, including frames, inversion, noise stability, exact alias identities, and two quantitative de-aliasing mechanisms. We then closed the main continuum bridge by proving that bounded gaps in active indices imply positive interior BM density with an explicit constant, converting gap regularization into an analytic injectivity hypothesis compatible with the MIT injectivity theorem.

Open Problem 12.1 (Certified soft-regularizer frame bounds beyond floors)**.** *Hard floors guarantee* $m_K(g) > 0$*. Can one design smooth regularizers whose minimizers imply quantitative lower bounds on* $m_K(g)$ and *on tail behavior, yielding certifiable conditioning without hard constraints?*

Open Problem 12.2 (Fast nonuniform algorithms)**.** *Extend the tight-frame theory to nonuniform grids (NUFFT regimes) and non-root-of-unity* β *sampling with certified condition numbers.*

Open Problem 12.3 (Learning nonseparable multivariate generators)**.** *Our* $\mathbb{T}^d$ *theory supports general multivariate coefficients* c_α*. Develop practical parameterizations of multivariate entire functions of finite order that are trainable at scale and compatible with BM-style density constraints.*

References

1. M. Abu-Ghuwaleh. *The Master Integral Transform with Entire Kernels*. *Mathematics* **13**(21):3431, 2025. DOI: 10.3390/math13213431.
2. A. Poltoratski. A problem on completeness of exponentials. *Annals of Mathematics* **178**(3), 983–1018, 2013.
3. A. Poltoratski. Spectral gaps for sets and measures. *Acta Mathematica* 208 (2012), no. 1, 151–209.
4. R. Giuliano and G. Grekos. On the connection between the Beurling–Malliavin density and the asymptotic density. *Uniform Distribution Theory* 19 (2024), no. 1, 67–86. doi:10.2478/udt-2024-0006.
5. A. Beurling and P. Malliavin. On the closure of characters and the zeros of entire functions. *Acta Mathematica* **118** (1967), 79–93. DOI: 10.1007/BF02392477.
6. P. Koosis. *The Logarithmic Integral, Volume I*. Cambridge University Press, 1988.
7. A. Rahimi and B. Recht. Random Features for Large-Scale Kernel Machines. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2007.
8. Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier Neural Operator for Parametric Partial Differential Equations. *arXiv:2010.08895*, 2020.
9. L. Dinh, D. Krueger, and Y. Bengio. NICE: Non-linear Independent Components Estimation. *arXiv:1410.8516*, 2014.

doi:10.20944/preprints202603.1171.v1

10. L. Dinh, J. Sohl-Dickstein, and S. Bengio. Density estimation using Real NVP. *arXiv:1605.08803*, 2016.
11. D. P. Kingma and P. Dhariwal. Glow: Generative Flow with Invertible 1×1 Convolutions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018. Also *arXiv:1807.03039*.
12. J.-H. Jacobsen, A. Smeulders, and E. Oyallon. i-RevNet: Deep Invertible Networks. In *International Conference on Learning Representations (ICLR)*, 2018. Also *arXiv:1802.07088*.
13. J. Behrmann, W. Grathwohl, R. T. Q. Chen, D. Duvenaud, and J.-H. Jacobsen. Invertible Residual Networks. In *International Conference on Machine Learning (ICML)*, 2019. Also *arXiv:1811.00995*.
14. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
15. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey,
ldots and others. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17 (2020), 261–272.
16. A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23) (2000), e215–e220.
17. S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2nd ed., 1999.